

International Journal of Modern Physics E
 © World Scientific Publishing Company

NUMERICAL CALCULATION WITH ARBITRARY PRECISION

BRUNO OSÓRIO RODRIGUES, L. A. C. P. DA MOTA and L. G. S. DUARTE

*Departamento de Física Teórica, Universidade do Estado do Rio de Janeiro, Rua São Francisco Xavier 524
 Rio de Janeiro, RJ 20550-900, Brazil
 brunooz@gmail.com*

Received (14 May 2007)

Revised (21 Jun 2007)

Accepted (21 Jun 2007)

The vast use of computers on scientific numerical computation makes the awareness of the limited precision that these machines are able to provide us an essential matter. A limited and insufficient precision allied to the truncation and rounding errors may induce the user to incorrect interpretation of his/hers answer. In this work, we have developed a computational package to minimize this kind of error by offering arbitrary precision numbers and calculation. This is very important in Physics where we can work with numbers too small and too big simultaneously.

1. Introduction

In the past, when doing calculations by hand, on paper, we imagined that the numbers we were using had the precision we wanted, i.e., virtually unlimited. When we wrote “2”, were actually regarding it as meaning “ $2.0 \cdots 0$ ”, with and infinity of zeros. Unfortunately, life is not that simple and the numbers in the computer can not have as much digits as desired. It is only able to represent numbers with a limited length. Today, it is usual to work with numbers which present 32bits precision, this allows for calculations with sixteen digit numbers. So, one can conclude that the computer is not able to represent real numbers. This limited precision of the computers may induce scientists to wrong answers, and worse: they may be using those wrong answers as if they were the right ones (see fig. 1) .

There are some ways to handle this problems, the obvious one is to use numbers as large as needed. In this undergraduate project, we are developing a computational package that permits us to make calculations with arbitrary precision numbers.

2. A Solution: NOz

NOz is the name of the computational package being developed^a in UERJ. It defines a new float point numerical type of arbitrary precision and the basic

^aBesides being developed, there is already a functional version of NOz

2 *B. O. Rodrigues, L. A. C. P. da Mota and L. G. S. Duarte*

$$\begin{array}{r} 560000 \\ -000010 \\ \hline 560000 \end{array}$$

Fig. 1. An example of a wrong answer if were used numbers with four digits of precision.

operations with them. Basically, it means that the user has the freedom to choose the precision of the numbers higher than the usual 32bits (or even smaller than it).

We know that there are already computational packages that can handle this problem, like GMP, but they are not easy to work with and demand some training before its use.^b Actually, the main feature that NOz has, as we shall explain later, is that it is very friendly. Just declare it in your main program and use it.

This package is written in pascal and in further versions, it will be compatible with C++ and Fortran programs.

3. The NOz Numerical Type

The NOz arbitrary precision type is a structure (see fig. 2) that contains the necessary informations of a float point number: signal (s), mantissa(M) and exponent(n).



Fig. 2. The structure of a real number.

In NOz, the signal is represented by a boolean type, the exponent is an integer type and the mantissa is a dynamic array of natural numbers. Each position of the mantissa array represents a digit of the number. It means that the mantissa array will have the same number of positions as the number of digits in the number. The positions are filled from the less significant digit to the most significant one. The complexity of the NOz algorithms is proportional to the number of positions in the mantissa. We are currently developing better algorithms in order to be able to use more than one digit per position in the mantissa array to improve the speed of the calculations.

^bGNU Multi-Precision - An open source package, written in C++ and assembler. More details in its web site <http://gmplib.org/>

The precision, in NOz, is not a property of each number in a calculation. It is defined as a variable in the main unit of the package and passed on as a parameter to the arbitrary precision functions. It is a way to ensure that, in an arithmetic operation, the numbers involved have the same precision. As it is a variable, the user has the freedom to change the precision of the calculations in runtime.

4. Numerical Operations Already Supported

It's already possible to use the following operations involving numbers of arbitrary precision using NOz: sum, subtraction, multiplication, division, comparison operations (equal to, minor than, greater than, etc) and the factorial.

With the four basic arithmetic operations, all the polynomial problems can be dealt with. Functions such as exponential, sin, cos and others can be treated by using series as the factorial function is already implemented.

5. Main Characteristics of NOz and Usage

The NOz (will) have the follow characteristics:

- Supports numbers with precision up to 2 billion digits.
- Easy to use.
- Allows the use of the operator symbols (+, −, * and /) by operator overload in the languages where it is possible, for instance, like Pascal.

To use NOz, the user has just to declare the `unoz.pas` unit in its main program, set the value of precision variable in `unoz` unit and ensure that the archive `noz.dll` (the compiled NOz package) is in the same directory of the main program. Code 1 is a little example of a program that calculates the exponential e^x using its Taylor series. The functions *sum*, *divis*, *multi* and *factorial* are the arbitrary precision functions for sum, division, multiplication and factorial present in NOz.

Using $x = 1$, $n = 70$ and the precision set to 100 digits, the program results:
 $e = 2.718281828459045235360287471352662497757247093699959574966967627724076630353547594571382178525166427E0.$

6. Conclusion

Being conscious of the limited precision problem is as important as being able to use arbitrary precision numbers. The NOz is a good solution to minimize the problem. Again, the problem is always there, we have to control it through using the NOz numbers that, although not being real numbers, present arbitrary precision, i.e., we can use the precision needed for the job at hand. The present version of NOz can deal with polynomial calculations only. But this is already of great use to the community since many interesting problems are polynomial in essence. Further versions will bring elementary functions as built-in operations. The main reason of its development is to give the community an easier way to handle the precision

4 B. O. Rodrigues, L. A. C. P. da Mota and L. G. S. Duarte

Code 1 Exponential

```

1  program Exponential_NOz;
2  {$APPTYPE CONSOLE}
3  uses
4      FastShareMem, SysUtils, UNOz {the NOz interface Unit};
5
6  function exponential (x : NOz_Float; n:integer) : NOz_Float;
7  var i : integer;
8      e_x : NOz_Float;
9  begin
10     e_x := NOz_0;
11
12     for i := 0 to n do
13         e_x := sum(e_x,divis(multi(x,x),factorial(i)));
14
15     result := e_x;
16 end;
17
18 var x : NOz_Float; // Declares x as an arbitrary precision float number
19     n : integer;
20 begin
21     x := StrToNoz('1.0E0'); //Conversion Routine from String to NOz_Float
22     n := 70; //Number of steps in the Exponential series
23
24     writeln(NozToStr(exponential(x,n))); //Converts result to String
25 end.
```

problem. But the use of arbitrary precision routines is indicated only when it is really necessary since, as we increase the precision of the calculation, we have a corresponding decreasing of the speed of the calculation.

Acknowledgements

Thanks to CNPq, FAPERJ and UERJ for the financial support and special thanks to Anibal L. Pereira, Luiz Fernando de Oliveira, J. Avellar and M. E. Bracco.

References

1. G.G. Langdon Jr, E. Fregni, *Projeto de Computadores Digitais* (Edgard Blücher, São Paulo, 1977)
2. John R. Hubbard, *Programming With C++* (McGraw-Hill, New York, 1996)
3. Behrooz Parhami, *Computer Arithmetic* (Oxford University Press, New York, 2000)
4. Marcus Cantù, *Dominando o Delphi 6*, “A Bíblia” (Makron Books, São Paulo, 2002)